

Normal Parameter Reduction Algorithm in Soft Set based on Hybrid Binary Particle Swarm and Biogeography Optimizer

Ali Safaa Sadiq^{1,2}, Mohammed Adam Tahir³, Abdulghani Ali Ahmed⁴, Abdullah Alghushami⁴

¹School of Mathematics and Computer Science,
University of Wolverhampton, WV1 1LY, UK
*ali.sadiq@wlv.ac.uk

²School of Information Technology, Monash University,
Bandar Sunway 47500, Malaysia

³Faculty of Technology Sciences, Zalingei University, Sudan

⁴Community College of Qatar,
P.O. Box 7344, Doha, State of Qatar

Abstract. Existing classification techniques that are proposed previously for eliminating data inconsistency could not achieve an efficient parameter reduction in soft set theory, which effects on the obtained decisions. Meanwhile, the computational cost made during combination generation process of soft sets could cause machine infinite state, which is known as Nondeterministic Polynomial time (NP). The contributions of this study are mainly focused on minimizing choices costs through adjusting the original classifications by decision partition order and enhancing the probability of searching domain space using a developed Markov chain model. Furthermore, this study introduces an efficient Soft-Set Reduction based Binary Particle Swarm optimized by Biogeography-Based Optimizer (SSR-BPSO-BBO) algorithm that generates an accurate decision for optimal and sub optimal choices. The results show that the decision partition order technique is performing better in parameter reduction up to 50%, while other algorithms could not obtain high reduction rates in some scenarios. In terms of accuracy, the proposed SSR-BPSO-BBO algorithm outperforms the other optimization algorithms in achieving high accuracy percentage of a given soft dataset. On the other hand, the proposed Markov chain model could significantly represent the robustness of our parameter reduction technique in obtaining the optimal decision and minimising the search domain.

Keywords: Classification, Markov chain model, Binary particle swarm optimization, Biogeography-based optimizer, Decision making.

1 Introduction

Nowadays, redundant data is one of the open issues due to the rapid development with technologies that inherently produce huge volume of data. This issue is more

visible especially in decision-making, since the behaviour of such type of data that gives more complexity and uncertainty during the process of decision making using such redundant data. Besides, the need of extra memory is essential which makes use of storage and produce redundant copies due to the widespread use. Thus, it has become a crucial need to reduce such huge amount of data which require substantial original soft data characteristics to improve the storage and resource utilization, hence a contribution is needed in improving the searching efficiency of an optimal decision for a given problem scenario. For this reason, the demands of reducing choices, cost, combinations complexity and memory space have encouraged researchers to develop smart techniques to address these issues of optioning optimal solutions, and at the same time these intelligent applications must inherit the characteristic of original soft data.

The issue of data redundancy and the use of its reduction have become a key challenge in knowledge management area. Considerable effort has been devoted in developing Knowledge Management Systems (KMS) to capture and manage knowledge through the digital capture, storage and retrievals not only in single location, but also in multiple distributions. KMS tried to overcome the replications problems of databases, which are geographically connected, to fit original sources in small space instead of data compressions (Min and Eom, 1994), (Maier, 2007), (Gottschalk, 2007), (Fulmer, 2011) and (Osei-Bryson et al. 2014). The success of KMS is that by its abilities in organizing and protecting the knowledge in an efficient manner (Min and Eom, 1994), that could help in better original characteristics and mining these data repositories to be managed more efficiently (Dalkir, 2013) and (Yu et al. 2009). The data redundancy and uncertainty information issues are formed and discussed by the emerged computing paradigm of information processing called Granular Computing (GrC) (Oscar Castillo and Pranab k. Muhuri, 2019), it emphasises on the need of developing a highly efficient data mining technique.

Simultaneously, data mining performance in discovering knowledge, which can delve into the massive data, can contribute to the knowledge discovery process, it relies on many factors, such as the data set size and the associated values (Chang et al. 2009). Taking into account the smaller amount of data should involve in knowledge processing stage such as extraction, reduction that helps various similarities techniques such clusters, rule associations, classifications to predict the result precisely and accurately. When the knowledge size is reduced, it is more easier to manage and enhance their generation, codification, transfer, access, store and analysis; however the process of selecting less coefficient data is critical and requires advances in information system (IS) (Merminod and Rowe, 2012; Marwan. et. al, 2018).

On the other hand, improving the cost performance of multi choices is important since for instance, in economic it assisting customers in the process of decision making. Besides, it can significantly save their money with help of obtaining optimal choice. This could be achieved when their taken process is rationalizing the cost and at the same time obtain an output has characteristics as exact as the original soft set. The data volume can be managed by IS. The IS facilitates data management by rapid processing as long as the key benefits clearly not only in the process of transmitting and exchanging information anywhere anytime, but their extensive use arise their

complexity and uncertain data. The complexity and uncertain data occur when information contain repetition or duplication, so it tends to be difficult for human to precisely understand their meanings, which could take days to solve such complex problems (Văduva, 2012; Ali Safa Sadiq. et. al, 2019). Therefore, good IS would produce more consistent decisions, which gives accurate result with low cost (Laudon and Laudon, 2009).

Normally, IS generates information with the help of Decision Support System (DSS) that distinguish between the cost of each choice by comparing their advantage and disadvantage. Using DSS as computer based model, IS able to solve the complex decisions problems significantly (Min and Eom, 1994). Besides, DSS can infer reasoning to organize and simplify knowledge management which describe the state of objects in discovering repeated data by using co-relations such as rules or constrains for long range plan decisions (Ayyub and Klir, 2010; Min and Eom, 1994). Through the DSS, mathematical models applied to computer system as a way to foster the calculations response time and with specific characteristics simplify the decision making process with prompt response (Min and Eom, 1994). However, these systems cannot make decision nor give reasoning, as they are normally used for verifying the data influences without precise and perfect theory, (Akerkar and Sajja, 2010). This yields limited results, and from this point the need of algorithm can overcome this issue to consistently achieving goals, is noticeable.

Many research studies have developed and viewed in favour of data reduction to ensure high quality and integrity of data before it can be processed in supporting decision-making. In the last decades several algorithms were developed in the field of data reduction which are aimed at obtain valuable information. Note that in case of big-data, the reduction process is very important since human beings' brains have limitation in performing decision-making. Human brain is only able to make decisions based on specific amount of information or choices, but cannot extend the decision beyond the provided information (Del Junco et al. 2010). As a result, the decision made by human brain is often not accurate in large data set, thus it is always subjected to uncertainly in the decision making process (Asemi et al. 2011), (Chen et al. 2009), (Min and Eom, 1994), (Babitha and Sunil, 2010) and (Ali Safa Sadiq. et. al, 2018).

As has been discussed in the preceding paragraphs, the existing methods of normal parameter reduction in soft set is still utilizing some kind of manual search, that is, by trial and error process, which could take more time. Furthermore, in case of large data in soft set, it is practically incredible to attain operative reduction. In this regards it is important mentioning this fact, the problem solving of normal parameter reduction is considered as one of the combinatorial problems. Besides, the nature-inspired methods includes swarm based techniques that mimic the social behaviour of groups of animals have been successfully applied in solving combinatorial optimization problems, which significantly save the time, money and manpower thus improve efficiency (Mirjalili, et. al, 2020). Hence, the proposed algorithm in this paper is significant to the body of knowledge, which could efficiently maintain the optimal and sub optimal choices during sot-set reduction process. Moreover, the new proposed Soft-Set Reduction based Binary Particle Swarm optimized by Biogeography-Based Optimizer

(SSR-BPSO-BBO) algorithm has efficiently addressed the issue of obtaining an optimal decision with high classification rate and less error values.

The reminder of this paper is structured as follows. In section 2, we will present the related works that are recently proposed in the area of study. Section 3 will introduce the preliminaries of soft set theory and definition, while the proposed normal parameter reduction algorithm will be detailed out in Section 4. Section 5 will discuss and illustrate on the performance analysis of our proposed algorithm, and followed by conclusion in section 6.

2 Related Work

Numerous research studies are proposed to deal with the complexity and uncertainty with data that inquire efficient techniques to reduce the data ambiguity for better decision-making. The nature of the uncertainty with data could be visible in many domains and it is varied from one to another. In this regards there are some of probability theories, such as fuzzy sets, rough sets, soft sets and other mathematical tools are famous and frequently used in handling such data uncertainty problems. Thus, it is obvious that as a way to enhance decision-making process, the uncertainty regions within data should be omitted. Hence, data reduction process should be taken place before performing any decision making process, which is assisting in obtaining optimal result. On the other hand, it is important mentioning that during data reduction process, it is very crucial to ensure that the obtained reduced sets still consist the original properties and attributes of the information (Zhao et al., 2007), (Rose et al., 2012). The main objective of reduction is to lessen the number of parameters, and at the same time, attaining the property of information in helping the process of decision-making (Rose et al., 2012; Mohammed Adam et al, 2017).

In data reduction research area, there are various techniques introduced by researchers such as (Maji et al., 2002), (Chen et al., 2005), (Kong et al., 2008), (Mamat, R. et al., 2011), (Rose et al., 2010), (Rose et al., 2012) and (Kumar and Rengasamy, 2013). Every soft set study has verified the influence of parameters exchanges during original combinations generating to search for exact decisions (solutions), therefore the direction of reduction measured by implicitly or non-implicitly conditions. The variation of implicitly or non- implicitly reductions noted that a multiples columns yield limited reduction results in the case of uncertainly in the form of non-implicitly. However, there are still issues and challenges in this research area, which give opportunity for further research to enhance the existing techniques. For example, reduction technique introduced by (Maji et al., 2002), has an issue of sub optimal problem which cause inconsistency in the obtained results. This problem of inconsistency from Maji was solved by (Chen et al., 2005), but the problem of sub optimal decisions though still existing, which induces incorrect and inconsistent obtained decision. Thus, to improve the accuracy of the decision making, (Kong et al., 2008) has introduced implicitly reduction technique. However, in Kong's

technique, in case there is no implicit representation of Normal Parameter Reduction (NPR), there will be no any reduction applied. Although, there is an implicit representation of NPR it has computational complexity, thus the reduction in (Kong et al., 2008) considered to be partially achieved. The complexity issue of proposed technique by Kong has been improved by (Rose et al., 2010) and (Mamat, R. et al., 2011). Furthermore, Rose in (Rose et al., 2012) has introduced a technique for identifying the soft set reduction base on implicitly representation, by merging all the NPR of a softset into a proposition of parameter Boolean variables. In another attempt, (Kumar and Rengasamy, 2013) has introduced another technique to reduce the soft data, but without considering implicit representation of the NPR. In Kumar's proposal, an argument was investigated on how to represent in a binary valued data the sample dataset at the same time the generated number of parameters is reduced without missing any of its original information.

Optimization algorithms are widely explored in the recent decades for finding the optimal solution in decision-making and data classification techniques. The optimization is such an arena, which can be applicable to attain an optimal solution containing discrete or continuous feasible solutions. Taking all things together, it can be stated that the general goal of either continuous or discrete optimization is to maximize or minimize a function. Alternatively, optimization is the economics of science and engineering with the fact of augmenting benefit, limiting expenses, industrial procedures or time utilization (Rubinovich and Miller, 2012). Various types of well-known optimizations are available in literature such as Combinatorial Optimization (Nemhauser and Wolsey, 2014), Complementarity Problems (Huang and Ni, 2010), Constrained Optimization (Bertsekas, 2014) and some more. However, this study has categorized these optimizations into two general categories such as discrete and continuous.

Minimizing or maximizing a function using continuous, real numbers by accepting value points from integer set to other set is known as Continuous optimization that contains negative values, decimals or fractions (Horst & Tuy, 2013). So, continuous optimization can take numerical values to make those values appear both in the real world and in the abstract mathematical world. Therefore, some experts believe that continuous optimization is more accurate and complex than its discrete counterpart (Streiner et al., 2014). However, many other experts oppose the finding (Seyedali Mirjalili. et. al, 2020b).

Conversely, a subclass of optimization, known as discrete optimization that can use integers as opposed to decimals or fractions and execute minimization or maximization of functions. Combinatorial optimization and integer programming are the two subdivisions of discrete optimization (Nemhauser & Bienstock, 2005). Precisely, the current study concentrates on developing Nature-inspired optimization algorithm that achieves the solution for continuous or discrete optimization problems stochastically.

In the previous couple of decades in scientific and engineering research, Nature-inspired algorithms are becoming progressively prevalent everywhere throughout the world. Researchers are getting excited by this improvement and have illustrated a few purposes behind this: portions of these causes are that they are created

to mimic the best elements in biological, chemical and physical processes in nature. This circumstance hurls the issue of deciding appropriate algorithm at whatever point a researcher has an optimization issue to solve. Usually, there is a common belief among the researchers that the decision of the 'best' algorithm to tackle a specific issue depends to a great extent on the kind of issue one is faced with. However, there are no such suggested guidelines on a decision of algorithm available for large-scale, non-linear optimization problems settling (Xu et al., 2012).

Meta-heuristic algorithms are prominent over few decades for solving difficult problems not only in computer science, but also for other fields since they are inspired by very simple natural selection concepts. Physical phenomena, animal behaviours and evolutionary concepts are the typical inspirations of meta-heuristic that facilitates the computer scientist to learn meta-heuristic, simulate various concepts, ensemble meta-heuristic with other algorithms, hybridize one with another, or improve existing meta-heuristic. Hence, the application of meta-heuristic algorithm to solve complex prediction problem consisting non-linear nature of data is a distinct research area that requires appropriate investigation. In a nutshell, meta-heuristic algorithms rely on two main components to perform the search process. Exploration is the process of roaming the entire search space to ensure sufficient diversity of the potential solutions. While exploiting is the process of finding the known best in the selected search region to ensure that the obtained solution is the most optimal. Excessive exploration tends to increase the computation and may lead to poor convergence. On the other hand, excessive exploitation can make the search process trapped in local optima without exploring other regions that might be more promising than the current one. For these reasons, there is a need to balance or tune between exploration and exploitation.

Given the aforementioned features, meta-heuristic algorithms can be applied for training neural network even though each algorithm has limitations. Some of the prominent meta-heuristic algorithms include Genetic Algorithm (GA) (Holland, 1992; and Goldberg, 1988), Particle Swarm Optimisation (PSO) (Kennedy, J. 2011), Bat Algorithm (BA) (Yang, X.-S, 2010a), Firefly Algorithm (FA) (Yang,X.S, 2010b), Cuckoo Search (CS) (Mohapatra. et. al, 2015), and Grey Wolf Optimiser (GWO) (Mirjalili, S., Mirjalili, S. M., & Lewis, A. 2014). However, no heuristic algorithm is the best suited to solve all optimization problems (Yang,X.S, 2012). Moreover, limitations of expensive computational cost, occurrence of premature convergence, mutation rate, crossover rate, time consuming fitness evaluation leads to enhance existing algorithm or propose new one. In machine learning, classification is a supervised learning process to determine appropriate dataset for a new observation based on the performance through training set. Evolutionary or nature-inspired algorithms are good option for classification. For example Support Vector Machine (SVM) and Artificial Neural Network (ANN) are efficient machine learning algorithms that are widely applied for classification. The optimization of the classifier parameters is possible through algorithms like GA, PSO, BAT, FA, and GWO, (Seyedali Mirjalili. et. al, 2020a). The parameter selection is a vital part of classification accuracy model and the parameter optimization of the classifier model through the application of meta-heuristic algorithms, which can simultaneously achieve the parameter selection. The parameter selection through this process is

another extension of distinct research dimension (Wei et al., 2017). However, these classifiers devise limitations such as: computationally expensive, high algorithmic complexity, extensive memory requirements, and selection of appropriate kernel parameters may be tricky (Sagar, 2015). Specifically, a problem well handled by a meta-heuristic may not produce same inspiring result for another problem.

As an attempt to develop metaheuristic algorithms, it is important to mention that, Simon has proven the Biogeography-Based Optimizer (BBO) algorithm is intelligent to overtake some well-known heuristic algorithms such as PSO, GA, ACO, ES, and Probability-based incremental learning (PBIL) (D. Dasgupta, Z. Michalewicz, 2001; I.C. Parmee, 2001) on fourteen benchmark functions and a real problem. He offered BBO as a competitive algorithm in the field of optimization (D. Simon, 2008).

The BBO algorithm was first proposed by Simon in (Simon, 2008). The basic concept BBO algorithm was inspired by biogeography, which relates to the study of biological organisms in terms of geographical distribution. In other words, Simon as way to introduce BBO algorithm adopted the basic concept of the geographical distribution over time and space. In BBO algorithm different islands, lands, or even continents over decades, centuries, or millennia could be included in a form of case studies. Thus, various ecosystems (habitats or territories) are explored to find the relationships between different species (habitants) in terms of immigration, emigration, and mutation. Hence, the essential inspiration for the BBO algorithm was come from the evolution of ecosystems in reaching a stable situation while considering different kinds of species (such as predator and prey), and the effects of migration and mutation. In the same way to other EAs, which are proposed by Wang (2012; 2013; 2014), BBO hires a number of search agents named habitats. These habitats are equivalent to chromosomes in GAs.

BBO is a meta-heuristic algorithm that applies evolutionary mechanisms to each individual in a population. BBO can provide more flexible training procedures compared to others for the search space of MLP that is changeable for different datasets. It tends to outperform GA due to applying various evolutionary operators (Mirjalili et al., 2014a). Usually, heuristic algorithms are employed for solving a particular problem by determining a combination of weights and biases that provide the minimum error for an MLP. The architecture does not change during the learning process in this method. For minimizing the overall error of MLP, the training algorithm needs to discover proper values for all connection weights and biases. Generally, there are three methods of using a heuristic algorithm for training MLPs. Firstly, heuristic algorithms are utilized for searching. Secondly, heuristic algorithms are employed to find a proper architecture for an MLP in a particular problem. The last method is to use a heuristic algorithm to tune the parameters of a gradient-based learning algorithm, such as the learning rate and momentum. The weights and biases are encoded using vector to train an MLP. The encoding is easier in this way though the decoding is a bit complicated. This method is used often for simple neural network structure and it is appropriate for the problem, which cannot deal with complex MLP structure (Haykin S., 1994).

BBO algorithm can be integrated to form ensemble approach for better classification and solving prediction problem. Moreover, BBO has much scope to grow, as this research community is quite young. Significant challenging tasks can be addressed through BBO by exploring new approach (Zhang et al., 2016, Mirjalili et al., 2014a). However, BBO algorithm may have some limitations such as poor in exploiting the solutions, no provision for selecting the best members from each generation and a habitat does not consider its resultant fitness while immigrating the features may result the generation of many infeasible solutions. The extension of BBO and ensemble with other models may be investigated to address the limitations (Ammu et al., 2013).

Regarding genetic algorithm, it maintains a population of possible solutions to the objective function being optimized. The early group of possible solutions is identified randomly. These possible solutions, called “chromosomes,” are permitted to grow over a number of generations. On every generation, the fitness of each chromosome is computed. This fitness will estimate the suitability of possible solution in optimizing the objective function. The preceding generation is shaped via a process of selection and recombination. The chromosomes are elected based on a probabilistic model for recombination based on their fitness. Throughout this process a measurement of how well the chromosomes in achieving the anticipated goal will be identified.

On the other hand, Ant Colony Optimization (ACO) (C. Blum, K. Socha, 2005) is one of the latest techniques for approximate optimization. The inspiring source of ACO algorithm is from real ant colonies. More precisely, ACO is inspired by the ants’ foraging behaviour. The indirect communication among ants is considered as the essential of foraging behaviour. This communication is conducted by means of chemical pheromone trails, which allows them to identify the short paths between their nest and food sources. This concept of real ant colonies is subjugated in ACO algorithm as a way to solve, for instance, discrete optimization problems.

Furthermore, in (S. Baluja, 1994) the population based incremental learning algorithm (PBIL) is proposed and considered as Estimation of Distribution Algorithms (EDA). PBIL assumes that all the variables of given problem scenario are independent. Besides, a probability vector is sustained during each phase of PBIL algorithm. This vector is represented by λ times to achieve λ new solutions. The $\mu \leq \lambda$ best solutions are elected and these have been used to modify the probability vector with a neural networks-inspired rule.

There is currently insignificant the literature focusing on the efficiency of BBO in training MLPs in the field of parameter reduction and decision-making based on soft-set technique. The only related work was by Kong (Kong, et. al, 2015). Kong has employed the particle swarm optimization algorithm to reduce the parameters. However, in the proposed algorithm the standard PSO was utilised, whereby standard PSO is lacking in terms of classification type of problems and does not assist in giving

a definite answer or one optimal decision (Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis, 2014; Mirjalili S., Amir H. Gandomi. 2017, Seyedali Mirjalili. et. al, 2020b). Thus, the binary version of PSO (BPSO) is more suitable with parameter reduction issue in soft-set, since it's dealing with two faces probability (0 or 1). Moreover, BPSO it needs an optimization process to take place for more accurate results. For this reason, some concepts of optimization process of MLP using BBO are discussed in this section.

To sum up the existing gaps within all aforementioned algorithms has motivated the researcher to develop a new algorithm that could efficiently identifies the optimal and sub optimal decisions through effective and improved reduction technique. Thus, in this research, there are two issues to be explored, which derived from the above issues in order to reduce the gaps. The first main issue need to be addressed is that reduce the data size while at the same time ensure the information is in a correct form, which keeps the main features of the original data. Afterwards, find the original features of the information through data classification of optimal and sub optimal decisions. The next section presents some preliminaries of the soft set theory.

3 Preliminaries

In this section, a definition of soft set is discussed in addition to an example presents the main concept of soft set theory for normal parameter reduction as well as its proposition.

Definition 3.1 (See (Maji et al. 2002). The authors employed the definitions of a pair (F, E) as a soft set over U , where F is mapping binary values from given parameters such $F : E \rightarrow P(U)$ including the generation of domain. For any original parameter sub set their relations are described based on its parameterized family as $\varepsilon \in E$, $F(\varepsilon)$ which overall is a subset of the universe U and has approximate elements can dealing with may, instead of a (crisp) set.

Example 3.1. As an illustration, consider the computational domain of a soft set (F, E) determined such as the “attractiveness of automotive promotions” which describes preference capabilities that for instance, Mr. X usually aims for when considering selections to purchase at less price and efficient preference. Assume the parameters in finite volume have thirty automotive promotion components in the universe U and the propositions are available under construction, $U = \{p_1, p_2, \dots, p_{30}\}$, and E was collocated as a set of choices (parameters). In sub set $E = \{e_1, e_2, e_3, e_4, p_5\}$ all variables were inserted then the concatenations interpret its meaning such as e_1 stored the values of “large tire” as first parameter, e_2 stored the values of “small tire” as second parameter, e_3 stored the values of “automatic” as third parameter, stored the

values of “manual” as four parameter, e_5 stands stored the values of “car status” as last parameter. The validations of this parameter mapping is derived as $F: E \rightarrow P(U)$ which described the influence of given “automotive promotions (\cdot) ”, where (\cdot) is used for choices to be filled in by complete decision. The decision from $e \in E$ categorized in levels as presented in Table 1.

Table 1. The mapping of parameters of example 3.1

$$\begin{aligned} F(e_1) &= \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}\}, \\ F(e_2) &= \{p_6, p_7, p_{18}\}, \\ F(e_3) &= \{p_5, p_6, p_7, p_8, p_{12}, p_{13}, p_{14}, p_{15}\} \\ F(e_4) &= \{p_3, p_5, p_6, p_9, p_{10}\} \\ F(e_5) &= \{p_1, p_5, p_6, p_7, p_8, p_{10}, p_{11}, p_{14}, p_{15}\} \end{aligned}$$

For example, $F(e_2)$ means automotive promotion for manual characteristic. The relationships of decision partitions were inferred among parameters as shown in Table 2. Table 2 shows relationship representations, which were established to help soft set algorithms for, drop up choices to significant choices and some choices may be negligible.

Thus, we can view and observe the soft set (F, E) as a collection of approximations categorized into similar classes as illustrated in the following:

Table 2. The decision partitions of soft set based on example 3.1

$$(F, E) = \left\{ \begin{aligned} e_1 &= \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}\} \\ e_2 &= \{p_6, p_7, p_{18}\} \\ e_3 &= \{p_5, p_6, p_7, p_8, p_{12}, p_{13}, p_{14}, p_{15}\} \\ e_4 &= \{p_3, p_5, p_6, p_9, p_{10}\} \\ e_5 &= \{p_1, p_5, p_6, p_7, p_8, p_{10}, p_{11}, p_{14}, p_{15}\} \end{aligned} \right\}$$

The information transfer into corresponding preferences is based on a correlation, where “1” will denote as part of the choice, and “0” means it is not part of choice of the automotive promotion as shown in Table 3.

Table 3. Tabular of soft set 1 parameters values from Example 3.1

U / E	p_1	p_2	p_3	p_4	p_5
u_1	1	0	0	0	1
u_2	1	0	0	0	0
u_3	1	0	0	1	0
u_4	1	0	0	0	0
u_5	1	0	1	1	1

u_6	1	1	1	1	1
u_7	1	1	1	1	1
u_8	1	0	1	0	1
u_9	1	0	0	1	0
u_{10}	1	0	0	1	1
u_{11}	1	0	0	1	1
u_{12}	1	0	0	1	0
u_{13}	0	0	1	1	0
u_{14}	1	0	1	1	1
u_{15}	1	0	1	0	1

The parameters governing flow are shown in Table 3, with the usage of “1” and “0” determining whether the objects can be part of preferences or not. In the table the parameter (preferences) flows can be dynamically managed by a Boolean-valued information system. The similar relations found in a soft set and a Boolean-valued information system is identified as follows:

Proposition 3.1. If (F, E) is a soft set the elementary over the universe U , then in (F, E) the calculations of a binary-value in information system is determined according to $S = (U, A, V_{\{0,1\}}, f)$ with obtained results similar to soft set correlation and gives result as shown in the previous tables.

Proof. Let (F, E) be a soft set over the universe U , then the $F = \{f_1, f_2, \dots, f_n\}$ mapping configuration is defined as $S = (U, A, V_{\{0,1\}}, f)$, where the evaluation of two terms of classification (binary) are assumed as $f_i : U \rightarrow V_i$ and $f_i(x) = \begin{cases} 1, & x \in F(e_i) \\ 0, & x \notin F(e_i) \end{cases}$, for $1 \leq i \leq |A|$ between any different configurations. Hence, if $A = E$, for any sub set is computed by $V = \bigcup_{e_i \in A} V_{e_i}$, where total parameters exchange $V_{e_i} = \{0,1\}$, then a soft set (F, E) correlations can be considered as a binary-valued information system .

From **Proposition 3.1** soft set revises the correlation between classes to make it easier and more understandable than a binary-valued information system approach and can be a suitable in representing and calculating every sub set of preferences from finite soft set. Thus, we can

make an judgement based on one-to-one correspondence between (F, E) over U and $S = (U, A, V_{[0,1]}, f)$.

This section is addressed the parameter mapping influences based on soft set theory. It has focused on the domain boundary of research problems namely the soft set exploration conducted in exact situation of previous research which did not completely fill the soft set reduction within the gap range as discussed in related works. However the soft set classification algorithm that proposed in this study is efficiently addressed this issue.

4 Proposed Normal Parameter Reduction Algorithm

In this section the detailed design of proposed normal parameter reduction algorithm based on adjusted weight vector and markov chain models is presented in utilising sub-section 4.1. Moreover, the intensive design of proposed Hybrid Binary Particle Swarm Optimization based on Biogeography-Based Optimizer algorithm (SSR-BPSO-BBO), is discussed with its role in supporting the classification process and decision making of our proposed normal parameter reduction algorithm in sub-section 4.2.

4.1 Proposed mathematical model of parameter reduction probability based on Adjusted Weight Vector and Markov Chain

The probability of occurrence of at least n suboptimal choice in a given soft-set can help us to investigate the key factors which have impact in suboptimal choice redundancy. In this regard, a suboptimal choice of subset P is considered as shown in Figure 1. Where in each choice combination in the subset P has a number of repetition and priority value.

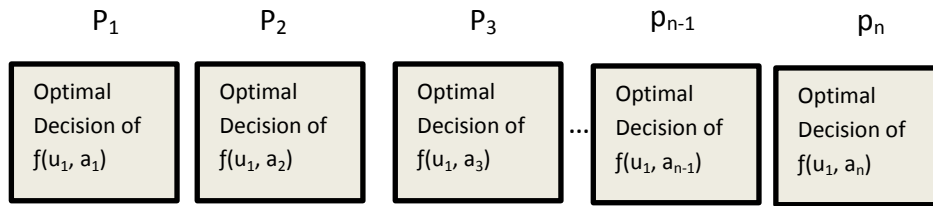


Figure 1. Probability representation of having an optimal decision within sub-set parameter of f_n

There are N sets of optimal decision in a soft-set and each combination of $f(u, a)$ has variables or features f_i and priority of reduction rate Pr . Optimal decisions are exponentially distributed in the soft-set with parameter λ , which refers to the mean value of P_i . Based on our assumptions, sub-sets are partially equipped with weighted priority which affects the optimal decision (Od) classified and ranked

according to the value of equation 1 based on the a proposed weight vector. W is identified to present the weight of Od.

$$\bar{W} = W_{f(u_1, a_1)}, W_{f(u_1, a_2)}, W_{f(u_1, a_3)}, \dots, W_{f(u_1, a_{n-1})}, \quad (1)$$

In order to obtain an optimal weight vector, an adjustment value is calculated with respect of standard deviation of each combination of suboptimal choice as presented by the following equation 2:

$$\begin{aligned} \bar{A} &= (A_{f(u_1, a_1)}, A_{f(u_1, a_2)}, \dots, A_{f(u_1, a_n)}) \\ &= \left(\frac{\sigma_{f(u_1, a_1)}}{i \sigma_i}, \frac{\sigma_{f(u_1, a_2)}}{i \sigma_i}, \dots, \frac{\sigma_{f(u_1, a_n)}}{i \sigma_i} \right) \quad i \in \{f(u_1, a_1), \dots, f(u_1, a_n)\} \end{aligned} \quad (2)$$

For instance, when the first sub-set PI of Od is considered to be given high priority based on the number of repetition, the given weight vector equation 3 will be as follows:

$$\bar{W} = \frac{A_{f(u_1, a_1)}}{M_{f(u_1, a_1)}}, M_{f(u_1, a_1)} \times A_{f(u_1, a_2)} \times A_{f(u_1, a_n)} \quad (3)$$

The standard deviation of the sub-set P values have been normalized in above equation, where the σ_i is the standard deviation of $P_{(i,1)}, P_{(i,2)}, \dots, P_{(i,n)}$ and M_i is their mean calculated utilizing the following equations 4 and 5:

$$M_i = \frac{1}{n}, n_{j=1} P_{i,j}, \quad i \in \{f(u_1, a_1), \dots, f(u_1, a_n)\} \quad (4)$$

$$\sigma_i = \frac{1}{n}, n_{j=1} (P_{i,j} - M_i)^2, \quad i \in \{f(u_1, a_1), \dots, f(u_1, a_n)\} \quad (5)$$

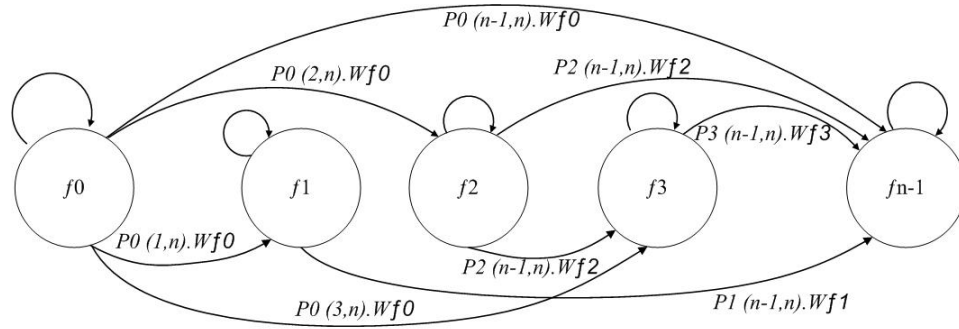
Based on our proposed method, the sub-set P with high probability of reduction is considered a main input metric for decision making in the optimal decision election process; thus, it should be highlighted that low variance of sub-set P should not be reflected in a decrease in its own weight vector. For example, when the overall average of $f(u_1, a_j)$, ($j = 1, 2, \dots, n$) is with high total number of active options/ones, the adaptation degree of sub-set metric $f(u_1, a_j)$ must be considered with high priority to be ranked as first optimal decision. Therefore, the weight vector of $f(u_1, a_j)$,

$w_{f(u_1, a_j)}$, should be given a higher weight value compare with the other optimal decision sub-sets. Accordingly, the reduction probability is significantly improved

by guaranteeing that the optimal decision was mad based on $f(u_1, a_j)$ with high priority weight vector.

Proposed Markov Model

Clearly, the parameter partition reduction probability of each parameter depends on the order of each partition in the sub-set. Hence, based on the order prosperities of parameters in the sub-set, it is likely that at least one or more sub-sets reduce with each cycle of obtaining an optimal decision. In order to compute the probability of at least x reduced sub-set in a cycle that called tail probability of x reduction and is shown with $Pr(x, n, m)$ in which n refers to the number of parameters with one sub-set and m is the possible number of combinations, we suggested an approach based on Markov modeling method. The proposed Markov model has N states in which each state f_i refers to at least i candidate parameter in the sub-set to be omitted from the new generated optimal decision set and also all transitions start from f_0 , Figure 2 presents the proposed Markov Chain model that demonstrating the probability concept of our parameterization value reduction technique. It is important to remind the readers that our proposed technique relying on three main phases, which are ranking the sub-sets f_n based on the number of parameters offering by each one (accumulative number of ones in a row), ranking each parameter by the number of its present with each f_n (accumulative number of ones in a column), and third phase is to guarantee an optimal decision is achieved by applying our proposed priority based weight vector.



$$[(0+Pr(1, n).Wf0) - (0+Pr(2, n).Wf0) - (0+Pr(3, n).Wf0) - \dots - (0+Pr(n-1, n).Wf0)], [f1], \dots, [fn-1]$$

Figure 2. Proposed Markov Chain model

For more understanding, an example is given of the way calculating the value of tail probability, when at least there is one sub-set selected to be reduced from the list of combinations.

$$P_r(1,4) = P_{f0} - P_{f1}(0 + P_{f0}) - P_{f2}(0 + P_{f0})(0 + P_{f1}) - P_{f3}(0 + P_{f0})(0 + P_{f1})(0 + P_{f2}) \quad (6)$$

P_r indicates to the probability of reduction rate of f_i in the sub-set, which is explained in the following. In aforementioned equation, the calculation of the tail probability of two parameters are reduced in a sub-set consist of five parameters, all possible combinations that could be used in selecting two options as an optimal decision out of five parameters should be considered. Assumed this theory, to calculate the tail probability of two reduced parameters in a sub-set, the tail probability of 1 parameter to be selected as the optimal decision should also been taken into account. Thus, transition between two consecutive states f_{i+1} and f_i where $i \geq 1$ is not possible and transition probabilities and state probabilities are same. The transition probabilities for at least one to three reduced parameters in N parameters length of a sub-set is shown as follows.

$$P_r(1, N) = \sum_{i=1}^{N-1} \left[P_{fi} \times \prod_{m=1}^{i-1} (0 + P_{fm}) \times W_{fm} \right] \quad (7)$$

$$P_r(2, N) = \sum_{i=1}^{N-2} \left[P_{fi} \times \left(\prod_{m=1}^{i-1} (0 + P_{fm}) \times W_{fm} \right) \times \left(\sum_{x=i+1}^{N-2} \left[P_{fx} \times \prod_{n=j+1}^{x-1} (0 + P_{fn}) \times W_{fn} \right] \right) \right] \quad (8)$$

$$P_r(3, N) = \sum_{i=1}^{N-3} \left[P_{fi} \times \left(\prod_{m=1}^{i-1} (0 + P_{fm}) \times W_{fm} \right) \times \left(\sum_{x=i+1}^{N-2} \left[P_{fx} \times \prod_{n=j+1}^{x-1} (0 + P_{fn}) \times W_{fn} \right] \right) \right. \\ \left. \times \left(\sum_{y=n+1}^{N-1} \left[P_{fy} \times \prod_{v=y+1}^{y-1} (0 + P_{fv}) \times W_{fv} \right] \right) \right] \quad (9)$$

Finally, we introduced a general recursive function for computing the tail probability of nr number of reduction in a sub-set which is represented by Equation 10 below:

$$P_r(f0, nr, N) = \left\{ \sum_{i=f0}^{N-nc} \left[P_{fi} \times \left(\prod_{j=f0}^{i-1} (0 + P_{fj}) \times W_{fm} \right) \times P(i+1, nr-1, N) \right], nr \neq 0 \right\} \quad (10)$$

$f0$ refers to the first parameter of the sub-set and N indicates the number of parameters given by a sub-set.

4.2 Proposed Soft Set Reduction using Hybrid Binary Particle Swarm Optimization based on Biogeography-Based Optimizer algorithm (SSR-BPSO-BBO)

Classification methods that are specialized to solve a specific problem can often achieve better performance in terms of accuracy and complexity time. This could be achieved by considering several features in addition to background knowledge should be obtained. In order to optimize the proposed Soft-Set-Reduction (SSR) algorithm, a new heuristic learning algorithm based on binary version of PSO (BPSO) is used to classify and reduce the low ranked parameters in a given soft-set. The binary version of this algorithm has been introduced for solving binary issues that representing with visibility of parameters within sub-sets of a universal soft-set, which are represented by 0s and 1s.

The BPSO was proposed by (Kennedy and Eberhartin 1997). Basically, the continuous and binary versions of PSO can be illustrated using two different components: a new transfer function and an altered technique for position updating process. For clarification, the process of mapping a continuous search space in a given soft-set to a binary one; a transfer function was used. On the other hand, the updating process is intended to switch positions of particles between 0 and 1 in binary search spaces.

In order to obtain an optimized classification method that could be integrated into SSR, Back Propagation (BP) of Multi-Layer Perceptron (MLP) Neural Networks (NNs) were used in this study. The new SSR using BPSO that is trained by Biogeography-Based Optimization (BBO) SSR-BPSO-BBO algorithm could efficiently assist in obtaining the optimal decision for the give problem scenario. The general mechanism which is used in generating the final decision of selected parameters of given sub-sets is designed to be BPSO. In order to modify particles position during searching process, in our proposed SSR-BPSO-BBO algorithm each particle within PSO should consider its current position, the present velocity, the distance to their personal best solution and the distance to the global best solution, gbest. The mathematical model of PSO is presented as follow:

$$V_i(t + 1) = rand \times V_i(t) + wc_i(t) \times c'_1 + c'_2 \times (gbest - X_i(t)) \quad (11)$$

$$X_i(t + 1) = X_i(t) + V_i(t + 1) \quad (12)$$

whereby $V_i(t)$ is the velocity of particle i at iteration t , w is a weighting function that was suggested earlier, c_j is an acceleration coefficient, $rand$ is a random number between 0 and 1, $X_i(t)$ is the current position of particle i at iteration t , and $gbest$ designates the best solution the swarm has achieved so far.

A random particle distribution in a problem space will be performed at the early stage of running a PSO. Afterwards, the velocities of particles are calculated using Eq. (11) and keep on updating during every iteration. When defining the velocities, the position of particles will be calculated using Eq. (12). The process of changing particles' positions will continue until satisfying an end criterion.

Commonly, there are many problems that have inherent distinct binary search spaces, like feature selection and dimensionality reduction (Mirjalili, Seyedali and Lewis, Andrew, 2013). Besides, issues with non-stop real-time search space can be converted into binary problems by converting their variables to binary variables, which is related to the behaviour of soft-set that normally consist of combinations of 0 and 1.

The main concept behind distinct binary searching spaces, the position updating process lays on switching between 0 and 1 value. Thus, this type of switching or position updating should be done based on the velocities of particles. Based on (S.Mirjalili, S.Z.Mohd Hashim, 2012), the concept of updating particles' velocity is introduced by applying a probabilistic model. The key idea is to change the position of any particle in a binary searching space with the probability of its velocity. Hence, to achieve this, a transfer function is necessary to change velocity values to probability values for updating the particle's positions.

$$T(v_i^m(t)) = \frac{1}{1 + e^{-v_i^m(t)}} \quad (13)$$

The velocity of particle i at iteration t in k -th dimension is $v_i^m(t)$. By changing velocities to probability values, position vectors could be updated with the probability of their velocities as follow:

$$x_i^m(t + 1) = \begin{cases} 0 & \text{if } rand < T(v_i^m(t + 1)) \\ 1 & \text{if } rand \geq T(v_i^m(t + 1)) \end{cases} \quad (14)$$

Figure 3 shows the steps of BPSO process in finding the optimal solution in a given searching space of soft-set parameter reduction. At the same time our proposed SSR-BBO algorithm it works in training the BP of MLP in obtaining the lowest Mean Square Error (MSE) of predicted output. Figure 4 illustrates the overall proposed SSR-BPSO-BBO flow activates. The elements with optimal MSE value will be used in updating the final position vectors of all particles in a searching space, which it finally helps in obtaining an optimal decision with high percentage of reduction of given soft-set.

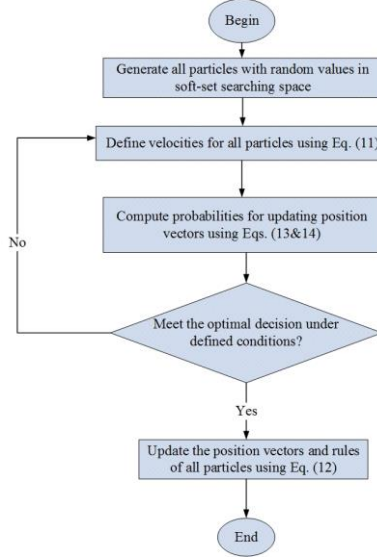


Figure 3. The flow chart of BPSO process

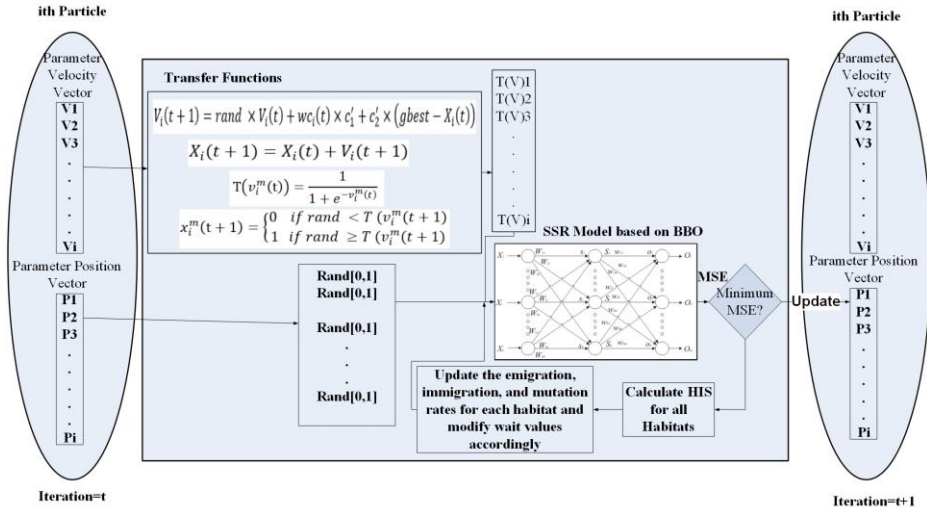


Figure 4. Proposed hybrid SSR-BPSO-BBO algorithm

The optimization of neural network can be performed through the application of Biogeography-Based Optimization (BBO) (Mirjalili et al., 2014a) in training MLPs. BBO is an evolutionary algorithm that applies evolutionary mechanisms to each individual in a population. BBO can provide more flexible training procedures

compared to others for the search space of MLP that is changeable for different datasets. It tends to outperform GA due to applying various evolutionary operators.

The pseudocode of BBO is as indicated in Figure 5. BBO algorithm will initially outline the island modification probability, mutation probability, and elitism parameter and initialize the population. The immigration rate and emigration rate will be calculated for each island, provided that the solution will be considered as good if it has high emigration rates and low immigration rates. Otherwise, if it has low emigration rates and high immigration rates then the solution will be treated as bad. Here, the immigration islands will be chosen based on the immigration rates probabilistically and roulette wheel selection will be used based on the emigration rates to select the emigrating islands. Then, randomly selected Suitability Index Variables (SIVs) will be migrated based on the selected islands where the migration will take place randomly. BBO performs mutation based on the mutation probability for each island probabilistically. Finally, fitness of each individual island will be calculated and the process continues until the target is achieved.

```

Begin
1: Generate an initial random population
2: While terminating condition is not reached do
3:   Calculate the immigration rate and emigration rate for each island
4:   Choose the immigration islands based on the immigration rates
   probabilistically
5:   Migrate randomly selected Suitability Index Variables (SIVs) based on the
   selected islands
6:   Perform mutation based on the mutation probability for each island
   Probabilistically
7:   Calculate the fitness of each individual island
8: End While
End

```

Figure 5. BBO Algorithm Pseudocode

In our proposed algorithm, the BBO sub-algorithm is applied to an MLP using the first method. Generally, the main aim of using BBO is train an MLP so that it is able to recognize training, validation, and testing the soft-sets completely for our use in parameter reduction process. The most significant sub-set in the learning phase is the training sub-set. The MSE function that was utilised in this study in indicating performance for individual sub-sets for all training samples is presented as follows:

$$E = \sum_{k=1}^n \frac{\sum_{i=1}^p (o_i^l - Opt_i^l)^2}{n} \quad (15)$$

n is the number of training samples, p is the number of outputs, Opt_i^1 is the optimal output of the i_{th} input unit when the k th training sample is used, and o_i^1 is the actual output of the i_{th} input unit when the k th training sample appears in the input parameters.

The BBO algorithm allocates each habitat a vector of habitants, which is similar to genes in a GA algorithm. This vector is representing the variables of problems. Moreover, using the defined Habitat Suitability Index (HSI) indicates the overall fitness of a habitat. Thus, higher value of HIS is representing the optimal habitat. Based on the concept of HIS, three main rules performed by habitats in order to progress over time as follows (Ma, 2013):

- Habitants living in habitats with high HSI are more likely to emigrate to habitats with low HSI.
- Habitats with low HSI are more prone to attract new immigrant habitants from those with high HSI.
- Habitats might face random changes in their habitants regardless of their HSI values.

BBO algorithm has utilized these concepts in developing the HSI of all habitats, which increases the number of initial random solutions for a particular problem scenario. Basically, BBO algorithm starts with a random set of habitats and every habitat has n different habitants correlated to the number of variables of an individual problem scenario. In addition, each habitat has its own immigration, emigration, and mutation rates. This simulates the typical geographically separated locations in nature.

Emigration α_k and immigration γ_k are expressed as functions of the number of habitants as follows:

$$\alpha_k = \frac{E \times n}{N} \quad (16)$$

$$\gamma_k = I \times \frac{1 - n}{N} \quad (17)$$

where n is the current number of habitants, N is the allowed maximum number of habitants, which is increased by HSI (the more suitable the habitat, the higher the number of habitants), E is the maximum rate of emigration, and I designates the maximum immigration rate.

Objective function

When the parameter set in a soft set be $\mathcal{F} = \{p_1, p_2, \dots, p_n\}$, and in order to find the parameter reduction is the way to find as minimal as necessary parameter subset in parameter set \mathcal{F} . Based on the feature of NPR in soft set, it is challenging to identify every parameter using real numbers. When using our proposed BPSO-BBO algorithm to perform the NPR, the searching agents, which are, binary strings will encode parameters. Presume that each parameter $u_i \in \mathcal{F}$ has a weight vector w'_i , and classifications of binary form are as follows:

if w'_i is 1, the matching parameter u_i will get selected; if w'_i is 0, the parameter u_i will be removed. Accordingly, the number of parameters in the soft set \mathcal{L} will be indicated by the length of the constricted binary string. Consequently, even though each particular parameter was selected can be represented using a series of n-bit binary. For instance, there is a parameter set $\mathcal{L} = \{p_1, p_2, \dots, p_8\}$, parameter subset $\{p_2, p_4, p_5, p_7\}$ is the normal parameter reduction of \mathcal{L} , then weight vector $W' = [w_1, w_2, \dots, w_8]$ can be represented by 01011010. Thus, as mentioned atelier, the NPR problem is to find as minimal as necessary parameter subset in parameter set \mathcal{L} . As a result, the objective function is defined as the minimal sum of W_i ($i = 1, \dots, n$). So, for soft set (S, \mathcal{L}) , $\mathcal{L} = \{p_1, p_2, \dots, p_n\}$, $\mu = \{u_1, u_2, \dots, u_n\}$, the objective function can be formulated as follows:

$$\begin{aligned} &\text{minimize } f(w) = \sum_{i=1}^n w_i \\ &\text{where } w = [w_1, w_2, \dots, w_n], w_i \in \{0,1\} \end{aligned}$$

Computational Complexity of SSR-BPSO-BBO:

The computational complexity of the SSR-BPSO-BBO depends on the number of generation (g), the population number (n), and the parameters dimensions (d). Therefore the overall computational complexity is $O(\text{SSR-BPSO-BBO}) = O(\text{Initialization}) + g(O(\text{Calculate the fitness of habitat}) + O(\text{Calculate the population in the Soft Set}) + O(\text{Sort the population based on BPSO}) + O(\text{Select n best habitat from the population}) + O(\text{Update the population}))$. The computational complexity of initialization is $O(nd)$, the computational complexity of calculating the archive population and fitness is $O(n)$, the computational complexity of sorting the population and archive population is $O(2n \log 2n)$, the computational complexity of selecting n best habitat from the population and opposition population is $O(n)$, the computational complexity of updating the population is $O(nd)$. Therefore, the final computational complexity is $O(\text{SSR-BPSO-BBO}) = O(nd) + g(2O(n) + O(2n \log 2n) + O(nd))$.

5 Performance Evaluation

In this section the performance evaluation of our proposed Normal Parameter Reduction Algorithm has been discussed and analysed in details. Moreover, a benchmark was conducted as a way to validate the performance of our proposed algorithm by fairly comparing it with other representatives in the field of normal parameter reduction in soft set. The following subsections are elaborating the performance evaluation for each stage of our proposed algorithm. Table 4 listed the simulation parameters of the implemented algorithms for NPR.

Table 4 Parameters for implemented Meta-Heuristic Algorithms

Algorithm	Parameter	Values
GA	Maximum number of generations	300
	Population size	100

	Type	Real coded
	Selection	Roulette wheel
	Crossover	Single point (probability = 1)
	Mutation	Uniform (probability = 0.01)
ACO	Maximum number of iterations	300
	Population size	100
	Initial pheromone (τ_0)	1e-06
	Pheromone update constant (Q)	20
	Pheromone constant (q_0)	1
	Global pheromone decay rate (p_g)	0.9
	Local pheromone decay rate (p_l)	0.5
	Pheromone sensitivity (α)	1
	Visibility sensitivity (β)	5
PSO	Maximum number of iterations	300
	Population size	100
	Topology	Fully connected
	Cognitive constant (C_1)	1
	Social constant (C_2)	1
	Inertia constant (w)	0.3
ES	Maximum number of iterations	300
	Population size	100
	λ	10
	δ	1

Table 4 Continue

PBIL	Maximum number of iterations	300
	Population size	100
	Learning rate	0.05
	Good population member	1
	Bad population member	0
	Elitism parameter	1
	Mutational probability	0.1
BPSO-BBO	Maximum number of iterations	300/600
	Population size	100
	Habitat modification probability	1
	Immigration probability bounds per gene	[0,1]
	Step size for numerical integration of probabilities	1
	Max immigration (I) and Max emigration (E)	1
	Mutation probability	0.005
	Number of particles	30
	c'_1	$\left(-\frac{2t^3}{T^3}\right) + 2$
	c'_2	$\left(\frac{2t^3}{T^3}\right)$
GWO	Maximum number of iterations	300

	Population size	100
	\vec{a}	Linearly decreased from 2 to 0
	\vec{A}	Random values in -2a to 2a
	\vec{C}	Random values in 0 to 2

5.1 Performance Validation of Proposed SSR-BPSO-BBO algorithm

In order to validate our proposed SSR-BPSO-BBO algorithm, the proposed Markov chain with respect to prioritized weighted vector is implemented using Matlab 2010 simulation tool, to prove numerically the performance of SSR. Figure 6 illustrating the achieved results that representing the probability of reduction using our proposed SSR-BPSO-BBO algorithm in the form of probability tail function and Probability Distribution Function (PDF). Figure 6a shows that our algorithm could achieve high fraction of total reduced parameters in probability tails of 0.9 when 16 parameters were given in a soft dataset. Meaning that our proposed SSR-BPSO-BBO has efficient ability in reducing parameters of soft dataset up to 90% (14.4 parameters were removed) out of 16 parameters.

On the other hand, Figure 6b demonstrates the PDF of our proposed SSR-BPSO-BBO algorithm with same number of parameters, (16) and random generated combinations of sub-set, (Number of combination 65534). It is obvious that based on our proposed Markov chain model, our SSR-BPSO-BBO algorithm could perform the best in terms of PDF in parameter reduction when the number of parameters and their random generated combinations are increased. Thus, a valid model was achieved by considering the weighted prioritized vector in the conducted decision making process of parameter reduction using proposed SSR-BPSO-BBO algorithm. The normal parameter samples are demonstrated in Figure 6c, that shows all generated samples and their value as well as covariance indexes. It is obvious that the data generated has difficult and complex behaviour that does not allow decision makers to obtain certain view, hence it proves the need for efficient NPR technique.

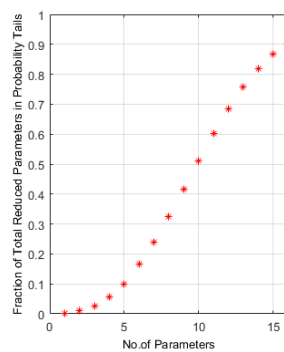


Figure 6a Fraction tail probability of proposed SSR-BPSO-BBO algorithm

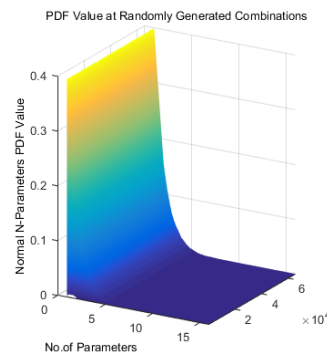


Figure 6b Normal N-Parameter PDF value

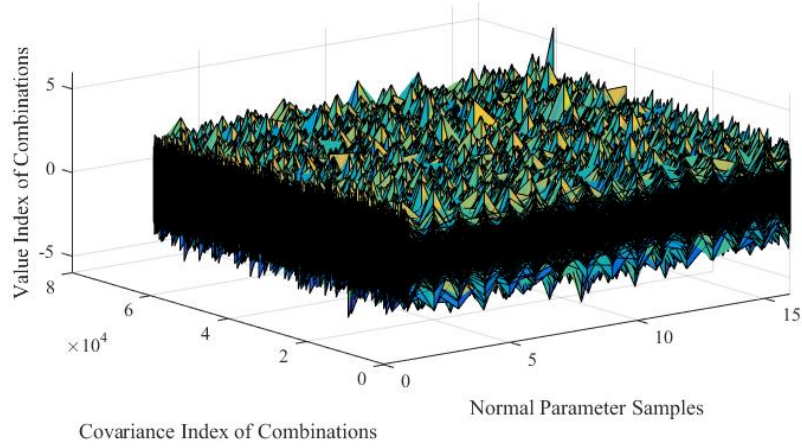


Figure 6c Normal parameter samples

Figure 6. Probability validation of proposed Markov chain model of SSR-BPSO-BBO algorithm

On the other hand, Figure 7 presents the modelled sequence of state transitions for parameter reduction using our proposed Markov chain model. We can observe that the state transitions of 15 parameters (15 states) and their probabilities of reduction P_r within the selected 32 different combinations sub sets. Utilising the aforementioned methods in Section 4, which are adjusted weight vector and priority; the P_r of 15 parameters are simulated and presented. While in Figure 8, a clear interpretation of performance analysis of our proposed Markov chain model was given. The probability to be selected as the best sub-optimal decision in a percentage form is presented for each individual parameter. We can read from the given graph in Figure 8 that parameters 10, 12 and 9 have obtained the highest probabilities respectively. In another words, parameters 10, 12 and 9 are the most highly recommended to remain as a sub-optimal decisions in the final reduced soft set. To remind the readers the reason behind this output is that, our proposed Markov chain model has considered the adjusted weight vectors and given priorities for each individual parameter in the soft set. The weight values were selected based in normalised weighted random numbers from 1 to 1000.

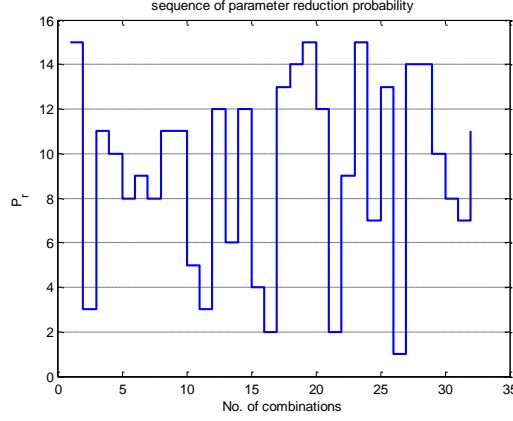


Figure 7. Sequence of state transitions for parameter reduction

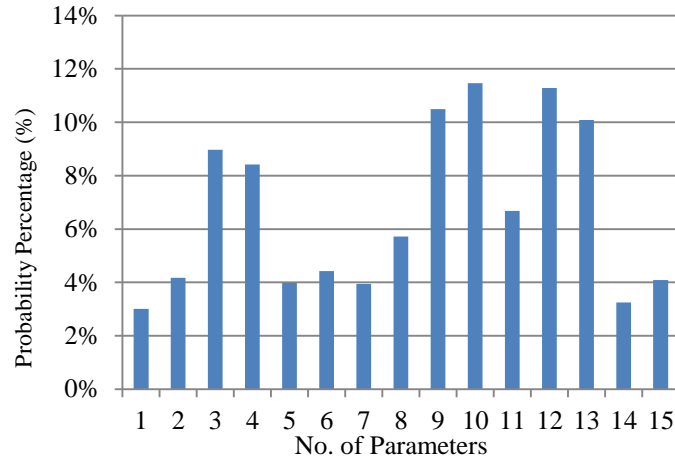


Figure 8. Probability percentage of the sub-optimal decision

5.2 Performance Evaluation of Proposed SSR-BPSO-BBO Algorithm

To evaluate our proposed SSR-BPSO-BBO algorithm a simulation was conducted based on Matlab 2010 tool. The binary version of PSO was implemented taking into account the recent improvements of particles' transfer functions V and S shaped that were proposed by Mirjalili, Seyedali and Lewis, Andrew (2013). Moreover, for more accuracy, the BBO optimization algorithm was implemented on generated soft dataset as a way to reduce the average error of decision making of our proposed SSR algorithm. Thus, the lowest obtained error out of learning process of conducted Neural Network (NN) is considered on updating the velocity and position of each particle

during searching process for the optimal decision. Hence, an optimized SSR algorithm was achieved in this study.

In Figure 9, the classification rates that were achieved by applying suggested BBO algorithm compared with other common optimization algorithms, (Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), Evolutionary Strategy (ES), Probability-Based Incremental Learning (PBIL) and Gary Wolf Optimizer (GWO)). It is important mentioning that each algorithm was run 10 times and the average percentage is presented in Figure 9. It is clear enough that the suggested BBO based optimizer has achieved higher overall classification rate of 89.33% compare with other optimization algorithms. While the second highest classification rate was achieved by GWO that is 87.50%. On the other hand, it was observed that ACO has obtained the lowest percentage that is 20% in classifying the given soft-set data. The reason behind that is BBO has the best ability in avoid trapping into local minima for the given dataset as its performance was dramatically improved with applying BPOS into its own process of tuning between exploration and exploitation. In another word, the modification that applied by the dynamic and random coefficient values could help the proposed algorithm in smoothly make the transition between exploration and exploitation trends with improved accuracy.

On the other hand, SSR based on BBO could achieve minimum overall MSE compare with other algorithms as illustrated in Figure 10 BBO tends to have the fastest convergence behaviour on provided dataset. The experiment was run for 10 times with 300 generations in each. The average MSE was calculated after each run and presented in Figure 10 as overall average MSE.

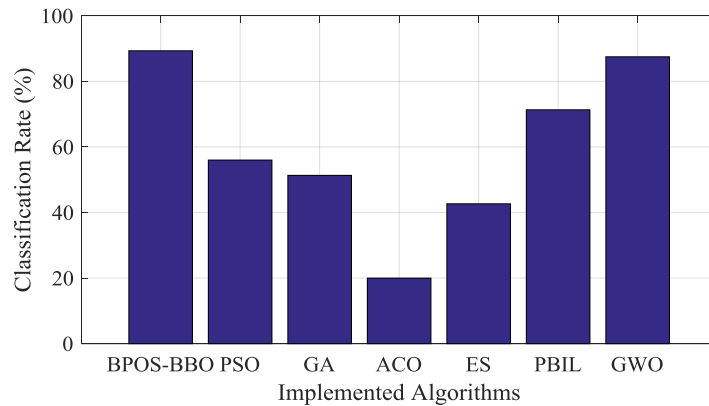


Figure 9. Classification rate of proposed SSR algorithm using BBO optimization algorithm compared with other optimization algorithms

Finally, the achieved results out of our proposed hybrid SSR-BPSO-BBO algorithm are presented in Figure 11. A soft dataset of Eight (8) sub-sets and 200 parameters ($v=200$) was generated randomly to test the ability of our proposed SSR-BPSO-BBO algorithm in selecting the optimal decision out of this big dataset during 600 general

iterations of BPOS with assistance of 300 iterations of BBO optimization algorithm. We predefined a threshold called Average-of-the-Best (AoB) and we identified the parameter reduction criteria to be $AoB \geq 50$. Hence, parameters achieve $AoB < 50$ will be omitted from the list of optimal decision.

Therefore, it is clear that sub-set number 4 has obtained the best convergence curves in average compared with others. The reason is that sub-set 4 could maintain a steady behaviour out of an average of 600 iterations. This was followed by sub-set 3 as a second ranked sub-set as it could achieve as lost similar average best-so-far compare with sub-set 4. While sub-set 7 has been removed since it does not satisfy our predefine constraint criteria. In other words, sub-set 7 obtained an AoB less than 50 during the last few iterations, thus it has been excluded from the list of optimal decisions. Hence, we could achieve a highly significant optimized soft-set reduction algorithm that could address weighty the issue of data redundancy compare with existing algorithms so far.

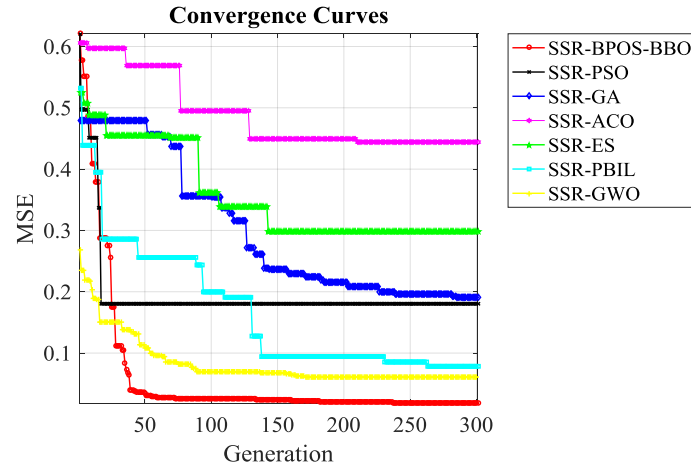


Figure 10. Convergence curves of MSE of SSR based BBO compared with other optimization algorithms of soft-dataset

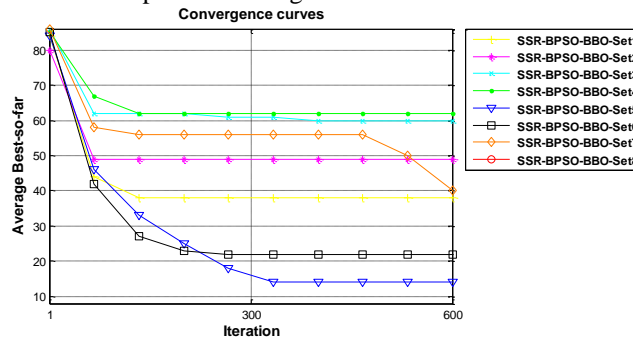


Figure 10. Convergence curves of Eight sub-sets of soft-dataset with $v=200$

6 Conclusion

In this paper two main contributions are proposed, which are the development of mathematical model that could efficiently formulate the probability of parameter reduction based on adjusted weighted vector and Markov Chain model, as well as maintaining soft set data classifications via the proposed SSR-BPSO-BBO algorithm which could utilise the soft-set data classification and obtained the optimal decision. In first component this study contributes to maintain consistency of optimal and sub optimal choices by the use of developed SSR-BPSO-BBO algorithm. The enhancement in the achieved cost of parameter reduction has benefited the customers and decision makers by utilising less data size, which could later on used through transmission lines in forecasting applications. The performance of proposed SSR-BPSO-BBO algorithm has solved the infinite machine state and improved the decision characteristics classifications by setting the decision partition and its order. The process of fixing original characteristics has produced minimized choice cost with high data quality, which has also simplified the decision representations. The reported results based on decision partition and its order have determined a new version of soft set reduction algorithm, which has uniquely addressed the issue of soft parameters reductions. On the other hand, the object or sub-set reduction has also been addressed in this paper. This was achieved via two stages, first by rotating two-dimensional tables using hybrid complement proposed algorithm SSR-BPSO-BBO to generate candidate objects solutions. Moreover, SSR-BPSO-BBO algorithm has enhanced the sub-set reduction by approximately 10% of improvement compared with the state of the art. On the other hand, the proposed SSR-BPSO-BBO algorithm could efficiently address the issue of choosing an optimal decision out of huge number of parameter combinations. This was achieved by considering the *S* and *V* shaped transfer functions based on binary version of PSO which was optimized via BBO to achieve low MSE of obtained decision. Thus, a significant contribution was achieved by obtaining a finite answer as the optimal decision.

In future studies, we will investigate the efficiency of our proposed algorithm by implementing it into other research fields such as wireless sensor networks and multimedia streaming, as well as forecasting and decision making problems.

Conflict of Interest

The author certifies that following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript.

REFERENCES

- Akerkar, R., and Sajja, P. 2010. Knowledge-based systems. Jones & Bartlett Publishers.
- Ali Safa Sadiq, Hossam Faris, Al-Zoubi Ala'M, Seyedali Mirjalili, Kayhan Zrar Ghafoor. 2019, Fraud Detection Model Based on Multi-Verse Features Extraction Approach for Smart City Applications, Smart Cities Cybersecurity and Privacy, pages:241-251, Elsevier.

- Ali Safaa Sadiq, Basem Alkazemi, Seyedali Mirjalili, Noraziah Ahmed, Suleman Khan, Ihsan Ali, Al-Sakib Khan Pathan, Kayhan Zrar Ghafoor. 2018. An efficient ids using hybrid magnetic swarm optimization in wanets, *IEEE Access*, Vol. 6, Pages:29041-29053.
- Asemi, A., Safari, A., and Zavareh, A.A. 2011. The role of management information system (MIS) and Decision support system (DSS) for manager's decision making process. *International Journal of Business and Management*. 6(7): p164.
- Ayyub, B.M., and Klir, G.J. 2010. Uncertainty modeling and analysis in engineering and the sciences. CRC Press.
- Babitha, K., and Sunil, J. 2010. Soft set relations and functions. *Computers & Mathematics with Applications*. 60(7): 1840-1849.
- Baluja S. 1994. "Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning," Technical Report CMU-CS-94-163(Computer Science Department, Carnegie Mellon University, Pittsburgh, 1994).
- Bertekas, D.P. (2014). *Constrained Optimization and Lagrange Multiplier Methods*: Academic Press.
- Blum C., Socha K. 2005. Training feed-forward neural networks with ant colony optimization: an application to pattern classification, in: *IEEE*, 6 pp.
- Chang, M.-Y., Hung, Y.-C., Yen, D.C., and Tseng, P.T. 2009. The research on the critical success factors of knowledge management and classification framework project in the Executive Yuan of Taiwan Government. *Expert Systems with Applications*. 36(3): 5376-5386.
- Chen, D., Tsang, E., Yeung, D.S., and Wang, X. 2005. The parameterization reduction of soft sets and its applications. *Computers & Mathematics with Applications*. 49(5): 757-763.
- Chen, Y.-C., Shang, R.-A., and Kao, C.-Y. 2009. The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electronic Commerce Research and Applications*. 8(1): 48-58.
- Dalkir, K. 2013. *Knowledge management in theory and practice*. Routledge.
- Dasgupta D., Michalewicz Z. 2001. *Evolutionary Algorithms in Engineering Applications*, Springer.
- Del Junco, J.G., Zaballa, R.D.R., and de Perea, J.G.Á. 2010. Evidence-based administration for decision making in the framework of knowledge strategic management. *Learning Organization*, The. 17(4): 343-363.
- Fine T.L. 1999. *Feedforward Neural Network Methodology*, Springer Verlag, 1999.
- Fulmer, Charles A. (2011). *Developing information storage and retrieval systems on the internet a knowledge management approach*. Monterey, California. Naval Postgraduate School.
- Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, 3(2), 95-99.
- Gottschalk, P. 2007. *Knowledge management systems in law enforcement: Technologies and techniques*. IGI Global.
- Guo Z.X., Wong W.K., Li M.2012. Sparsely connected neural network-based time series forecasting, *Inform. Sci.* 193 (2012) 54-71. ISSN 0020-0255, <http://dx.doi.org/10.1016/j.ins.2012.01.011> (15.06.12).
- Herawan, T., Rose, A.N.M., and Deris, M.M. 2009 a. Soft set theoretic approach for dimensionality reduction Database Theory and Application (pp. 171-178): Springer.
- Herawan, T. 2014. Recent Advances on Soft Computing and Data Mining: Proceedings of the First International Conference on Soft Computing and Data Mining (Scdm-2014) Universiti Tun Hussein Onn Malaysia, Johor, Malaysia June 16Th-18Th 2014. Springer.
- Herawan, T., Deris M.M. 2009 b. A direct proof of every rough set is a soft set.
- Herawan, T., Deris M.M. 2009 c. A soft set approach for association rules mining. *Knowledge- Based System*. 24(1), 186-195, 2011.
- Holland, J. H. (1992). Genetic algorithms. *Scientific American*, 267(1), 66-73.
- Horst, R., & Tuy, H. (2013). *Global Optimization: Deterministic Approaches*: Springer Science & Business Media.
- Huang, Z.-H., & Ni, T. (2010). Smoothing Algorithms for Complementarity Problems over Symmetric Cones. *Computational Optimization and Applications*, 45(3), 557-579.
- Ji, M., Han, J., and Danilevsky, M. (2011). Ranking-based classification of heterogeneous information networks. Paper presented at the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.
- Kennedy J., Eberhart R. 1997. A discrete binary version of the particle swarm algorithm, in: *Proceedings of the IEEE International Conference on Computational Cybernetics and Simulation*, 1997.
- Kennedy, J. (2011). Particle swarm optimization. In *Encyclopedia of machine learning* (pp.60-766). Boston: Springer.

- King W.R. 2009. Knowledge Management and Organizational Learning. Annals of Information Systems 4, LLC 2009.
- Kong, Z., Gao, L., Wang, L., and Li, S. 2008. The normal parameter reduction of soft sets and its algorithm. Computers & Mathematics with Applications. 56(12): 3029-3037.
- Kong, Z, Jia, W, Zhang, G and Wang, L. 2015. Normal parameter reduction in soft set based on particle swarm optimization algorithm, journal of Applied Mathematical Modelling, vol 39, no. 16, pp: 4808—4820, 2015.
- Kumar, D.A., and Rengasamy, R. 2013. Parameterization reduction using soft set theory for better decision making. Paper presented at the Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference.
- Laudon, K., and Laudon, J. 2009. Management Information Systems: International Edition, 11/E. Pearson Higher Education.
- Maier, R. 2007. Knowledge management systems: Information and communication technologies for knowledge management. Springer.
- Maji, P., Roy, A.R., and Biswas, R. 2002. An application of soft sets in a decision making problem. Computers & Mathematics with Applications. 44(8): 1077-1083.
- Mamat, R., Herawan, T., and Deris, M.M. 2011. Super attribute representative for decision attribute selection Software Engineering and Computer Systems (pp. 137-147): Springer.
- Marwan Batrouni, Aurélie Bertaux and Christophe Nicolle. 2018. Scenario analysis, from BigData to black swan. Computer Science Review. Volume 28, Pages 131-139. Doi: 10.1016/j.cosrev.2018.02.001.
- Merminod, V., and Rowe, F. 2012. How does PLM technology support knowledge transfer and translation in new product development? Transparency and boundary spanners in an international context. Information and Organization. 22(4): 295-322.
- Miller, B. M., & Rubinovich, E. Y. (2012). Impulsive Control in Continuous and Discrete-Continuous Systems: Springer Science & Business Media.
- Mirjalili S., Amir H. Gandomi. 2017 Chaotic gravitational constants for the gravitational search algorithm, Applied Soft Computing, vol. 53, pp: 407–419.
- Mirjalili S., Hashim S.Z.M. 2010. A new hybrid PSO-GSA algorithm for function optimization, in: 2010 International Conference on Computer and Information Application (ICCIA), 3–5 December 2010, pp. 374, 377.
- Mirjalili S., Lewis, Andrew and Ali Safa Sadiq 2014 Autonomous particles groups for particle swarm optimization, Arabian Journal for Science and Engineering, V39, No. 6, PP. 4683—4697, 2014.
- Mirjalili S., Mirjalili S.M., and Andrew Lewis. 2014 Let a biogeography-based optimizer train your multi-layer perceptron, Information Sciences, vol. 269, pp: 188—209.
- Mirjalili S., Mohd Hashim S.Z., BMOA 2012. binary magnetic optimization algorithm, International Journal of Machine Learning and Computing, vol 2, no. 3, PP: 204–208, 2012.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014b). Grey Wolf Optimizer. Advances in Engineering Software, 69, 46-61.
- Mirjalili, Seyedali and Lewis, Andrew 2013. S-shaped versus V-shaped transfer functions for binary particle swarm optimization, Swarm and Evolutionary Computation, Vol. 9, pp:1-14, 2013.
- Mohammed Adam Taheir Mohammed, Ali Safa Sadiq, Ruzaini Abdullah Arshah, Ferda Ernawan, Seyedali Mirjalili. 2017. Soft Set Decision/Forecasting System Based on Hybrid Parameter Reduction Algorithm, Journal of Telecommunication, Electronic and Computer Engineering (JTEC), Vol. 9, No. 2-7, Pages:143-148.
- Mohapatra,P.,Chakravarty,S.,&Dash,P.K.(2015).An improved cuckoo search based extreme learning machine for medical data classification. Swarm and Evolutionary Computation, 24, 25–49.
- Nemhauser, G., & Bienstock, D. (2005). Integer Programming and Combinatorial Optimization: Springer.
- Oscar Castillo, Pranab k. Muhuri, Special issue on “Type-2 fuzzy systems and granular computing”, Granular Computing, Vol 4, Issue 2, pp 143-143, April 2019
- Osei-Bryson, K.-M., Mansingh, G., and Rao, L. 2014. Knowledge Management for Development: Domains, Strategies and Technologies for Developing Countries. Springer Science & Business Media.
- Parmee I.C. 2001. Evolutionary and Adaptive Computing in Engineering Design, Springer Verlag, 2001.
- Rose, A.N.M., Herawan, T., and Deris, M.M. 2010. A framework of decision making based on maximal supported sets Advances in Neural Networks-ISBN 2010 (pp. 473-482): Springer.
- Rose, A.N.Mohd, Awang, M.I., Hassan, H., Zakaria, A. H., Herawan, T., and Deris, M.M. (2012). Hybrid reduction in soft set decision making Advanced Intelligent Computing (pp. 108-115): Springer.
- Sagar S. Nika. (2015). A Comparative Study of Classification Techniques in Data Mining Algorithms. Oriental Journal of Computer Science & Technology, 8(1), 13-19.

- Seyedali Mirjalili, Jin Song Dong, Ali Safa Sadiq, Hossam Faris. 2020a, Genetic Algorithm: Theory, Literature Review, and Application in Image Reconstruction, Nature-Inspired Optimizers, pages:69-85, Springer.
- Seyedali Mirjalili, Jin Song Dong, Andrew Lewis, Ali Safa Sadiq. 2020b, Particle Swarm Optimization: Theory, Literature Review, and Application in Airfoil Design, Nature-Inspired Optimizers, pages:167-184, Springer.
- Simon D. 2008. Biogeography-based optimization, IEEE Trans. Evol. Comput. 12 (2008) 702–713.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2014). Health Measurement Scales: A Practical Guide to Their Development and Use: Oxford University Press.
- Văduva, I. 2012. On Solving some types of Multiple Attribute Decision Making Problems. Romanian Journal of Economic Forecasting. 15(1): 41-61.
- Wang G.-G., Gandomi A.H., Alavi A.H. 2013 A chaotic particle-swarm krill herd algorithm for global numerical optimization, Kybernetes 42 (6) 962–978.
- Wang G.-G., Gandomi A.H., Alavi A.H. 2013. An effective krill herd algorithm with migration operator in biogeography-based optimization, Appl. Math. Model, in press, <http://dx.doi.org/10.1016/j.apm.2013.10.052>.
- Wang G.-G., Gandomi A.H., Alavi A.H. 2014. Stud krill herd algorithm, Neurocomputing 128 (2014) 363–370, <http://dx.doi.org/10.1016/j.neucom.2013.08.031>.
- Wang G.-G., Gandomi A.H., Alavi A.H., Hao G.-S. 2013. Hybrid krill herd algorithm with differential evolution for global numerical optimization, Neural Comput. Appl, <http://dx.doi.org/10.1007/s00521-013-1485-9>.
- Wang G., Guo L., Wang H., Duan H., Liu L., Li J. 2012. Incorporating mutation scheme into krill herd algorithm for global numerical optimization, Neural Comput. Appl, <http://dx.doi.org/10.1007/s00521-012-1304-8>.
- Wei, Y., Ni, N., Liu, D., Chen, H., Wang, M., Li, Q. & Ye, H. (2017). An Improved Grey Wolf Optimization Strategy Enhanced SVM and Its Application in Predicting the Second Major. Mathematical Problems in Engineering, 2017(1), 1-12.
- Wolsey, L.A., & Nemhauser, G. L. (2014). Integer and Combinatorial Optimization: John Wiley & Sons.
- Xu, H., Caramanis, C., & Mannor, S. (2012). Sparse Algorithms Are Not Stable: A No-Free-Lunch Theorem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(1), 187-193.
- Yang, X.-S. (2010a). A New Metaheuristic Bat-Inspired Algorithm Nature Inspired Cooperative Strategies for Optimization (NICSO 2010) 65-74: Springer.
- Yang, X.-S. (2012). Swarm-Based Metaheuristic Algorithms and No-Free-Lunch Theorems: Theory and New Applications of Swarm Intelligence.
- Yang, X.S. (2010b). Firefly algorithm, stochastic test functions and design optimisation. International Journal of Bio-Inspired Computation, 2(2), 78–84.
- Yu, H., Huang, X., Hu, X., and Wan, C. 2009. Knowledge management in E-commerce: A data mining perspective. Paper presented at the Management of e-Commerce and e-Government, 2009. ICMECG'09. International Conference on.
- Zhang N. 2009. An online gradient method with momentum for two-layer feedforward neural networks, Appl. Math. Comput. 212 (2009) 488–498.
- Zhao, Y., Luo, F., Wong, S.M., and Yao, Y. 2007. A general definition of an attribute reduct Rough Sets and Knowledge Technology (pp. 101-108): Springer.